Supplementary Information

Recognition of methylated DNA through methyl-CpG binding domain proteins

Xueqing Zou[1], Wen Ma[1,2], Ilia A. Solov'yov[1], Christophe Chipot[1,3], and Klaus Schulten*[1,2,4]

[1]Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana, Illinois, USA

[2]Center for Biophysics and Computational Biology, University of Illinois, Urbana, Illinois, USA

[3]Equipe de Dynamique des Assemblages Membranaires, UMR Centre National de la Recherche Scientifique/UHP 7565, Nancy Universite BP 239, Nancy, France

[4]Department of Physics, University of Illinois, Urbana, Illinois, USA

October 25, 2011

*To whom correspondence should be addressed. Email: kschulte@ks.uiuc.edu

Supporting Tables

time (ns)	$E_{\mathrm{m}}^{(1)}$ $(kcal/mol)$	$E_{\mathrm{n}}^{(1)}$ $(kcal/mol)$	$E_{\mathrm{m}}^{(1)} - E_{\mathrm{n}}^{(1)}$ $(kcal/mol)$
20	-41.88	-41.17	-0.71
21	-41.67	-41.49	-0.18
22	-39.5	-38.65	-0.85
23	-41.96	-41.62	-0.34
24	-43.38	-42.65	-0.73
25	-39.12	-37.83	-1.29
26	-42.49	-41.68	-0.81
27	-42.26	-41.05	-1.21
28	-42.02	-41.06	-0.96
29	-40.23	-39.36	-0.87
30	-43.05	-42.09	-0.96
time (ns)	$E_{\mathrm{m}}^{(2)}$ $(kcal/mol)$	$E_{\mathrm{n}}^{(2)}$ $(kcal/mol)$	$E_{\mathrm{m}}^{(2)} - E_{\mathrm{n}}^{(2)}$ $(kcal/mol)$
20	-43.22	-42.55	-0.67
21	-42.91	-42.41	-0.5
22	-39.97	-39.39	-0.58
23	-41.03	-40.91	-0.12
24	-40.21	-39.76	-0.45
25	-22.51	-21.32	-1.19
26	-42.19	-41.22	-0.97
27	-41.20	-40.25	-0.95
28	-40.64	-38.90	-1.74
29	-41.14	-40.60	-0.54
30	-40.93	-39.85	-1.08

Table S1: Quantum chemistry calculations for three body interaction energies in stair motifs. $E_{\mathrm{m}}^{(1)}$ is the three body interaction energy for mCYT106 $\therefore$ ARG22 $\vee$ GUA107, and $E_{\mathrm{n}}^{(1)}$ nCYT106 $\therefore$ ARG22 $\vee$ GUA107; $E_{\mathrm{m}}^{(2)}$ is the interaction energy for mCYT118 $\therefore$ ARG44 $\vee$ GUA119, and $E_{\mathrm{n}}^{(2)}$ nCYT118 $\therefore$ ARG44 $\vee$ GUA119. The results account for correction of the base set superposition error (see text).

Supporting Figures S1 - S6

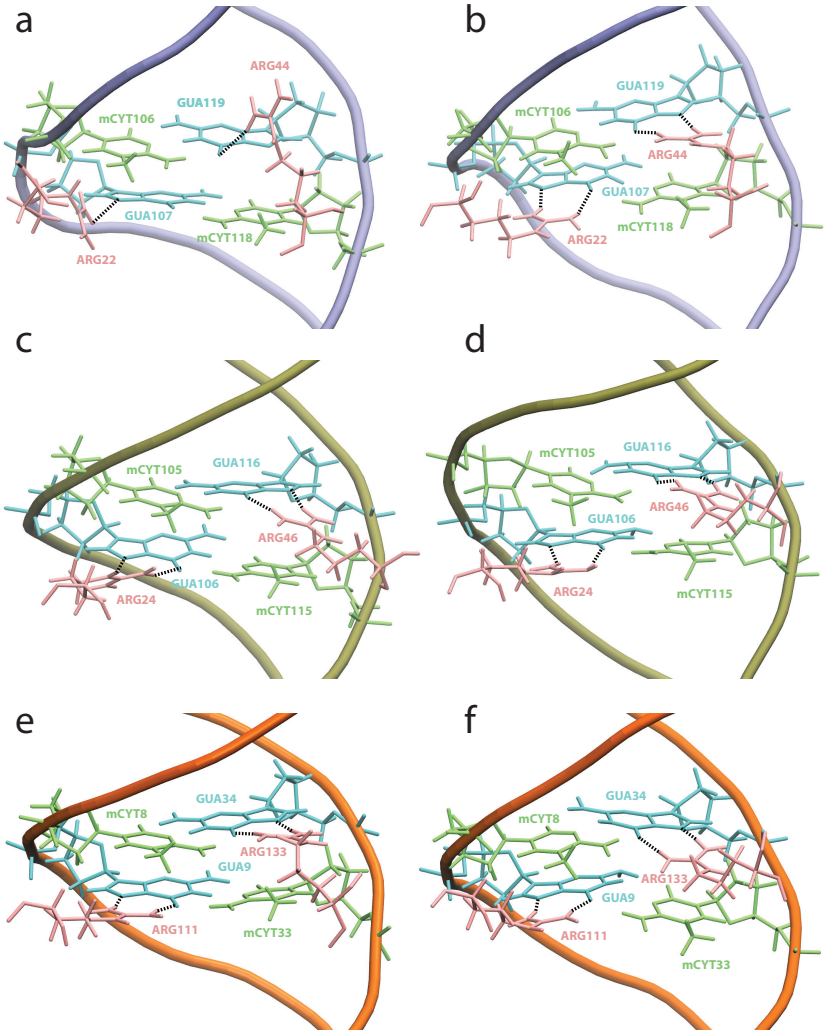

Figure S1: Comparison between crystallographic structures and MD simulation structure of stair motifs in MBD1-mDNA, MBD2-mDNA and MeCP2-mDNA complexes. (a) NMR structure of the MBD1-mDNA complex. (b) Structure of the MBD1-mDNA complex after 30 ns simulation. (c) NMR structure of the MBD2-mDNA complex. (d) Structure of the MBD2-mDNA complex after 30-ns simulation. (e) X-ray structure of the MeCP2-mDNA complex. (f) Structure of the MeCP2-mDNA complex after 30 ns simulation.

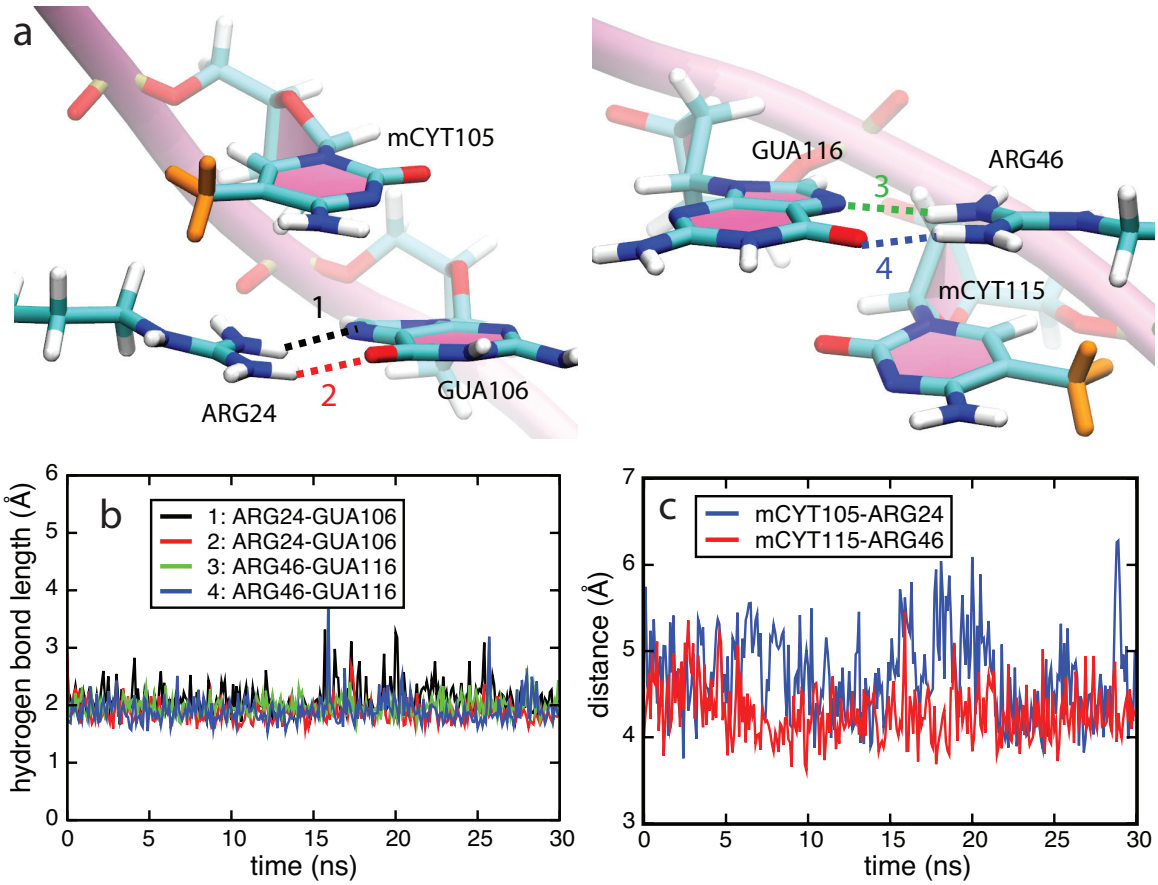

Figure S2: Stair motifs in the interface of MBD2-mDNA. (a) Orientation of stair motifs mCYT105 ∴ ARG24 ∨ GUA106 (left) and mCYT115 ∴ ARG46 ∨ GUA116 (right) in the MBD2 protein. mCYT indicates the methylated cytosine residues. (b) Time evolution of hydrogen bonding between H- and N-atoms in the ARG24-GUA106 pair (black line), H- and O-atoms in the ARG24-GUA106 pair (red line), H- and N-atoms in the ARG46-GUA116 pair (green line), H- and O-atoms in the ARG46-GUA116 pair (blue line). (c) Time evolution of the distance between the center of mass of mCYT105 residue in the DNA and ARG24 residue in the MBD2 protein (blue line), and the distance between the center of mass of mCYT115 residue in the DNA and ARG46 residue in the MBD2 protein (red line).

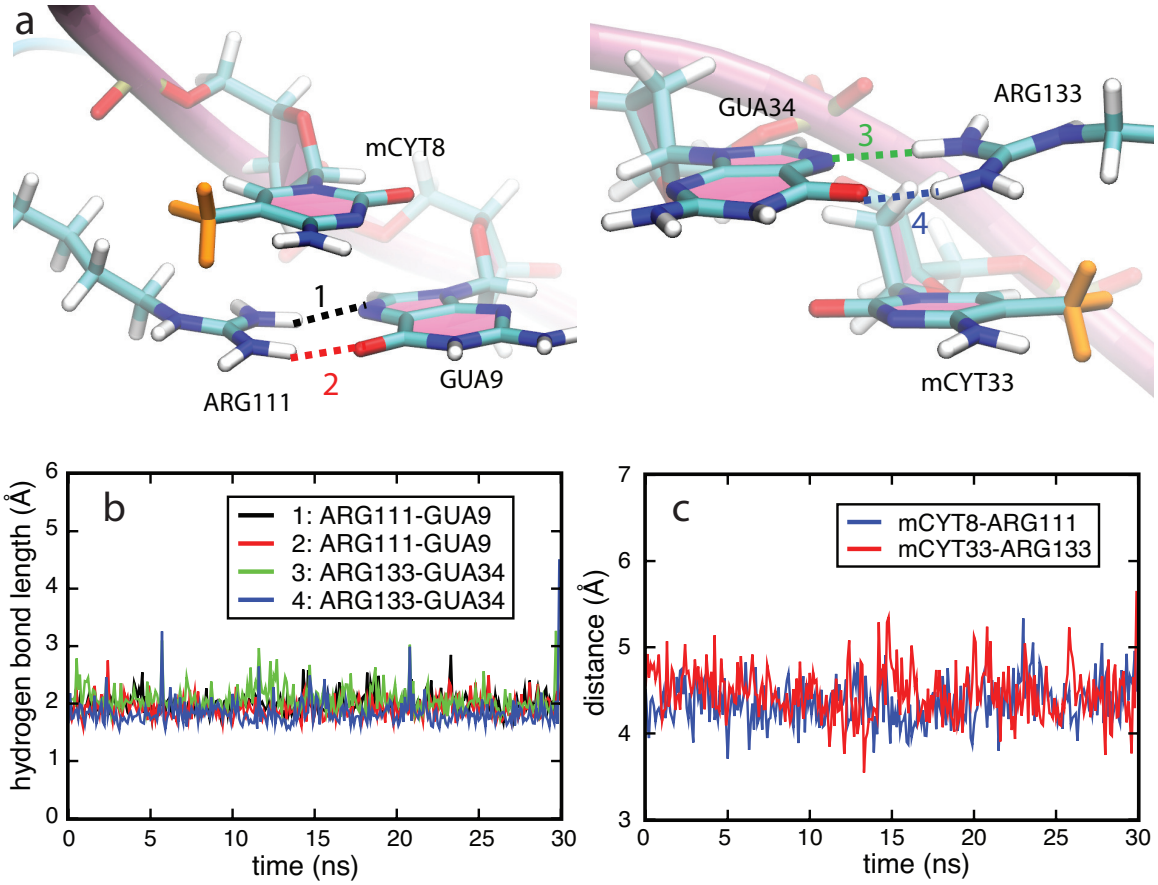

Figure S3: Stair motifs in the interface of MeCP2-mDNA. (a) Orientation of stair motifs mCYT8 ∴ ARG111 ∨ GUA9 (left) and mCYT33 ∴ ARG133 ∨ GUA34 (right) in the MeCP2 protein. mCYT indicates the methylated cytosine residues. (b) Time evolution of hydrogen bonding between H- and N-atoms in the ARG111-GUA9 pair (black line), H- and O-atoms in the ARG111-GUA9 pair (red line), H- and N-atoms in the ARG133-GUA34 pair (green line), H- and O-atoms in the ARG133-GUA34 pair (blue line). (c) Time evolution of the distance between the center of mass of mCYT8 residue in the DNA and ARG111 residue in the MeCP2 protein (blue line), and the distance between the center of mass of mCYT33 residue in the DNA and ARG133 residue in the MeCP2 protein (red line).

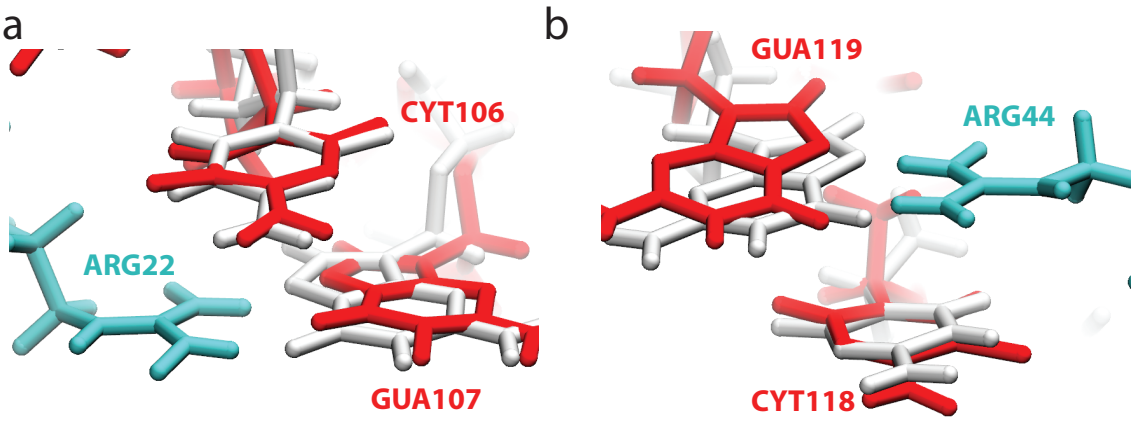

Figure S4: Displacement of guanine due to arginine binding. (a) mCYT106 ∴ ARG22 ∨ GUA107 stair motif in the MBD1-mDNA complex. (b) mCYT118 ∴ ARG44 ∨ GUA119 stair motif in the MBD1-mDNA complex. Nucleotides of B-form DNA are shown in grey. After aligning cytosines, the GUA107 and GUA119 of MBD1-mDNA complex show a displacement towards to the minor groove of DNA, reducing the stacking area with its neighboring cytosine.

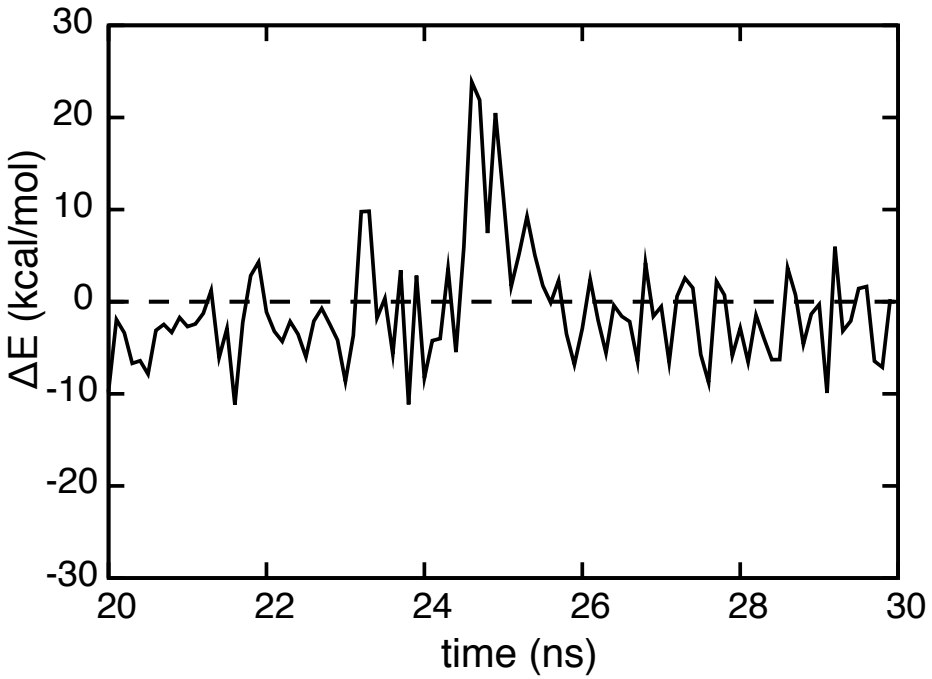

Figure S5: Influence of cytosine methylation on stair motif binding energy in MBD1-DNA complex. The energy difference ΔE is calculated based on the CHARMM force field, showing the methylation-induced three-body interaction energy difference of the mCYT $\therefore$ ARG $\vee$ GUA stair motifs on the MBD1-DNA interface.

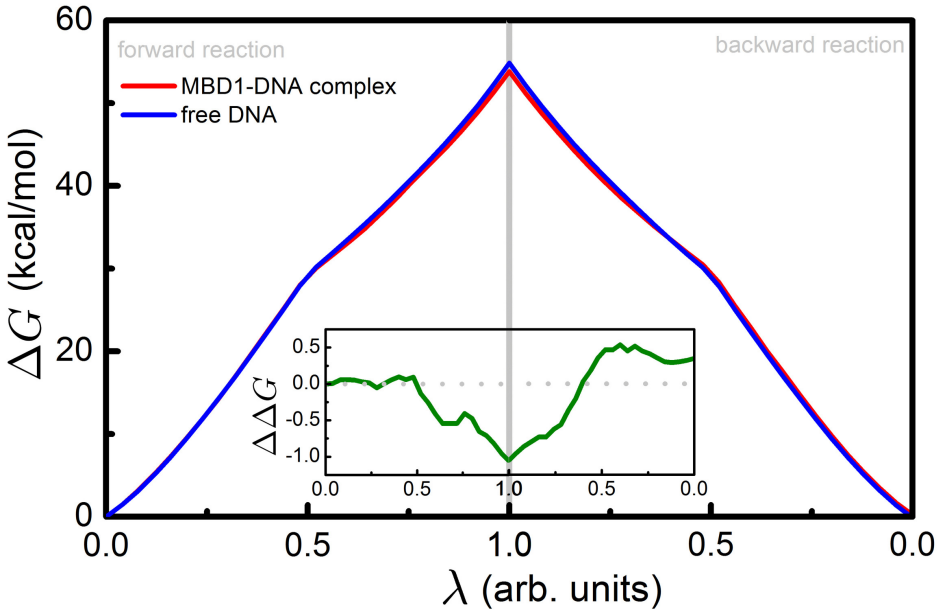

Figure S6: Free-energy change for methylating two cytosines (CYT106 and CYT118) in free DNA (blue) and in MBD1-DNA complex (red). The net free-energy change for MBD-DNA $\rightarrow$ MBD-mDNA due to methylation is shown in the inset (green). The ΔG obtained from simulations of forward (left) and backward (right) transformations in water are shown.

Supporting Movies

- **Movies S1 - S3** show the structures of MBD1-mDNA (**S1**), MBD2-mDNA (**S2**) and MeCP2-mDNA (**S3**) complexes. See also Fig. 2 for static image of the complexes. The methylated cytosines in DNA are indicated in red.

- **Movies S4 - S5** show trajectories of the mCYT $\therefore$ ARG $\vee$ GUA (**S4**, simulation (i) in Table 1) and CYT $\therefore$ ARG $\vee$ GUA (**S5**, simulation (ii) in Table 1) stair motifs.

Supporting Results

Quantum chemistry calculation for the interaction energy between CYT106 and ARG22

Ab initio quantum chemistry calculation of the interaction energies were performed using Møller-Plesset perturbation theory (MP2) (1–4), coupled cluster single and double excitation (CCSD) expansion (5) and density functional theory (DFT) (6–9). These methods were used to calculate the interaction energy between CYT106 and ARG22 of the MBD1-DNA complex extracted at equal time intervals from molecular dynamics (MD) simulations of methylated and non-methylated DNA in complex with MBD1. Prior to the quantum chemistry calculations, the geometry of each MBD1-DNA complex taken from MD simulations, was optimized to the nearest local minimum employing the CHARMM27 force field (10). The methylated state of the CYT-ARG fragment contains 35 atoms. It includes a methylated cytosine and an arginine, along with two hydrogen atoms terminating the side-chains of these residues. Naturally, the non-methylated state of the CYT-ARG fragment contains 32 atoms. The two-body interaction energy between the CYT106 and ARG22 is defined as

$$E_{\rm m} = E({\rm mCYT-ARG}) - E({\rm mCYT}) - E({\rm ARG}) \quad \text{(S1)}$$

$$E_{\rm n} = E({\rm nCYT-ARG}) - E({\rm nCYT}) - E({\rm ARG}). \quad \text{(S2)}$$

Here the subscripts m and n denote methylated and the non-methylated forms of the CYT106-ARG22 pair, respectively. $E({\rm CYT-ARG})$ is the total energy of the CYT106-ARG22 pair, while $E({\rm CYT})$ and $E({\rm ARG})$ are the energies of the individual CYT106 and ARG22 residues.

The change of the two body interaction energy due to DNA methylation can be calculated as

$$\Delta E = E_{\rm m} - E_{\rm n}. \quad \text{(S3)}$$

The MP2 calculations were performed in conjunction with the 6-311++G(d, p) and cc-pVDZ basis sets used for the electronic wave function expansion. The latter basis set is the so-called correlation consistent basis set of double-zeta quality (11). The interaction energies

for CYT106 and ARG22 are given in Table S2. These energies are in reasonable agreement with the result of calculations of cytosine and arginine interactions in other protein-DNA complexes (12).

The results of the CCSD calculations, which are considered to be more accurate than the MP2 method, are compared with the results of the latter method. Though the absolute values for the interaction energies are different, the energy differences between the methylated and the non-methylated states show a good agreement between CCSD and MP2 calculations (see Fig. S7).

The difference between the results obtained within the framework of the DFT method and other methods demonstrates the important role of the van der Waals interactions of dispersive nature on the binding energy of the CYT106 and ARG22 (see Fig. S7).

MP2/6-311++G(d, p)			
time (ns)	E_{m} ($kcal/mol$)	E_{n} ($kcal/mol$)	$E_{\mathrm{m}} - E_{\mathrm{n}}$ ($kcal/mol$)
25	0.78	1.66	-0.88
26	0.29	0.33	-0.04
27	0.34	0.44	-0.10
28	0.77	0.85	-0.08
29	-0.53	0.21	-0.74
30	1.17	1.29	-0.12
MP2/cc-pVDZ			
time (ns)	E_{m} ($kcal/mol$)	E_{n} ($kcal/mol$)	$E_{\mathrm{m}} - E_{\mathrm{n}}$ ($kcal/mol$)
25	1.11	1.95	-0.84
26	0.36	0.43	-0.07
27	0.41	0.51	-0.10
28	0.78	0.89	-0.11
29	-0.34	0.33	-0.67
30	1.16	1.35	-0.19
CCSD/cc-pVDZ			
time (ns)	E_{m} ($kcal/mol$)	E_{n} ($kcal/mol$)	$E_{\mathrm{m}} - E_{\mathrm{n}}$ ($kcal/mol$)
25	2.23	2.87	-0.64
26	1.18	1.01	0.17
27	1.19	1.20	-0.01
28	1.30	1.42	-0.12
29	0.63	1.12	-0.49
30	1.89	1.97	-0.08
B3LYP/6-311++G(d, p)			
time (ns)	E_{m} ($kcal/mol$)	E_{n} ($kcal/mol$)	$E_{\mathrm{m}} - E_{\mathrm{n}}$ ($kcal/mol$)
25	5.26	5.14	0.12
26	3.40	2.27	1.13
27	3.18	2.71	0.47
28	2.53	2.54	-0.01
29	3.22	2.92	0.30
30	3.86	3.34	0.52

Table S2: Quantum chemistry calculation for two body interaction energy of CYT106 and ARG22. E_{m} is the interaction energy for mCYT106 and ARG22, and E_{n} nCYT106 and ARG22. The title row for each block states the method and the basis set used in the calculations (e.g. MP2/6-311++G(d, p) corresponds to the results obtained using the second-order Møller-Plesset perturbation theory in conjunction with the 6-311++G(d, p) Pople-type basis set). The results take into account correction for the base set superposition error (15).

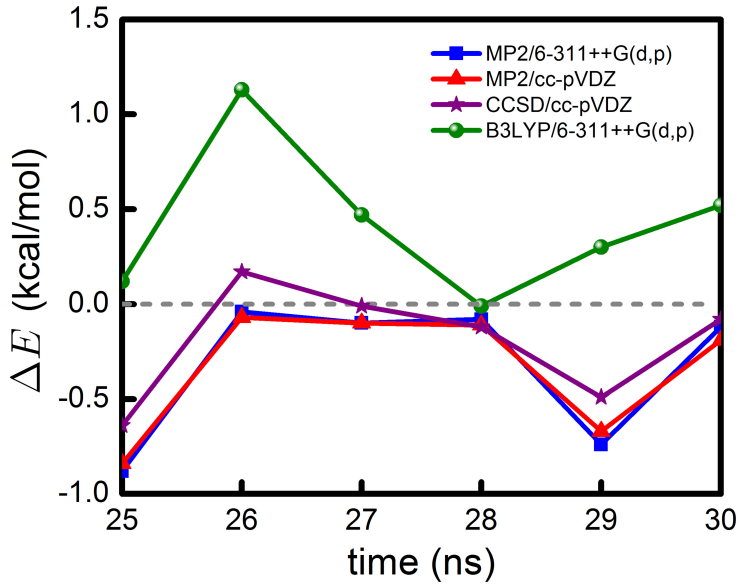

Figure S7: Effect of cytosine methylation on the interaction energy between CYT106 and ARG22. The energy difference shown is calculated with different quantum-chemistry methods. The time evolution corresponds to six snapshots taken from MD simulation trajectory in the 25–30 ns interval. Four different sets of results corresponding to different methods or basis sets are shown (see text for more details).

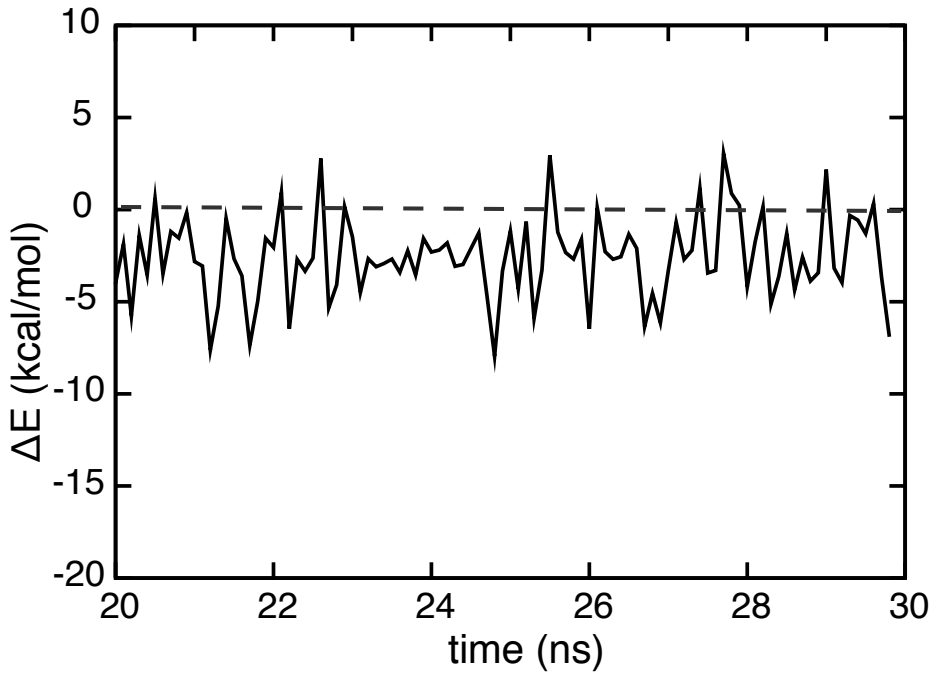

Figure S8: Effect of cytosine methylation on the interaction energy between CYT106 and ARG22. The energy difference shown is calculated based on the CHARMM force field. The change of the two body interaction energy due to DNA methylation can be calculated as: $\Delta E = E_{\mathrm{m}} - E_{\mathrm{n}}$. Here $\Delta E < 0$ shows a stabilization effect upon cytosine methylation, which is consistent with QM calculations.

References

1. Møller, C. and Plesset, M. S. (1934) Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.,* **46**, 618–622.

2. Foresman, J. B. and Frisch, A. (1996) Exploring Chemistry with Electronic Structure Methods, Gaussian Inc., Pittsburgh, PA.

3. Wintjens, R., Biot, C., Rooman, M., and Liévin, J. (2003) Basis Set and Electron Correlation Effects on ab Initio Calculations of Cation-π/H-Bond Stair Motifs. *J. Phys. Chem. A,* **107**, 6249–6258.

4. Biot, C., Wintjens, R., and Rooman, M. (2004) Stair motifs at protein-DNA interfaces: nonadditivity of H-Bond, stacking, and cation-π Interactions. *J. Am. Chem. Soc.,* **126**, 6220–6221.

5. Piecuch, P., Kucharski, S. A., Kowalski, K., and Musial, M. (2002) Efficient computer implementation of the renormalized coupled-cluster methods: The R-CCSD[T], R-CCSD(T), CR-CCSD[T], and CR-CCSD(T) approaches. *Comput. Phys. Commun.,* **149**, 71–96.

6. Parr, R. G. and Yang, W. (1989) Density-functional theory of atoms and molecules, Oxford University Press, New York.

7. Yakubovich, A., Solov'yov, I., Solov'yov, A., and Greiner, W. (2006) Conformational changes in glycine tri- and hexapeptide. *Eur. Phys. J. D,* **39**, 23–34.

8. Solov'yov, I., Yakubovich, A., Solov'yov, A., and Greiner, W. (2006) Ab initio study of alanine polypeptide chain twisting. *Phys. Rev. E,* **73**, 021916–(1–10).

9. Solov'yov, I., Yakubovich, A., Solov'yov, A., and Greiner, W. (2006) On the fragmentation of biomolecules: fragmentation of alanine dipeptide along the polypeptide chain. *J. Exp. Theor. Phys.,* **103**, 463–471.

10. MacKerell, Jr., A., Bashford, D., Bellott, M., Dunbrack, Jr., R. L., Evanseck, J., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, I. W. E., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins.. *J. Phys. Chem. B,* **102**, 3586–3616.

11. Dunning, T. H. (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.,* **90**, 1007–1023.

12. Wintjens, R., Liévin, J., Rooman, M., and Buisine, E. (2000) Contribution of cation-π interactions to the stability of protein-DNA complexes. *J. Mol. Biol.,* **302**, 393–408.

13. van Mourik, T. and Gdanitz, R. J. (2002) A critical note on density functional theory studies on rare-gas dimers. *J. Chem. Phys.,* **116**, 9620–9623.

14. Šponer, J., Jurečka, P., and Hobza, P. (2006) Base Stacking and Base Pairing. In Šponer, J. and Lankaš, F., (eds.), *Computational Studies of RNA and DNA*, pp. 343–388 Springer Netherlands.

15. Boys, S. F. and Bernardi, F. (1970) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.,* **19**, 553–566.